

# Digitising Handwritten Parish Records using Deep Learning

SoDa Labs - Monash Business School

# How does the data look like?

## Overview

- Birth, baptism, marriage, and death records from Austria (and Germany)
  - Period: 1625-1938
  - 183,000 church registers from Austria; ~4.5 million images

A church register from Austria, 1855 (source: <https://data.matricula-online.eu/>)

# Aim

***To digitise tabular data contained in raw images (unstructured data) in the form of spreadsheet (structured data)***

2 sub-aims:

- Layout analysis
- Optical character recognition (OCR)

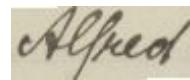
# Aim: Layout Analysis

To discern structure of tables found in parish registers images:

- Identify field names
- Find separator lines (for segregating rows and columns)
- Crop snippets of individual text-entities to be used in OCR
- Position each text-entity into its corresponding row and column

# Aim: Optical character recognition (OCR)

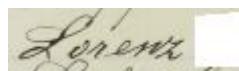
To parse handwritten text imprinted on the images to produce machine-readable text:



Alfred



Maria



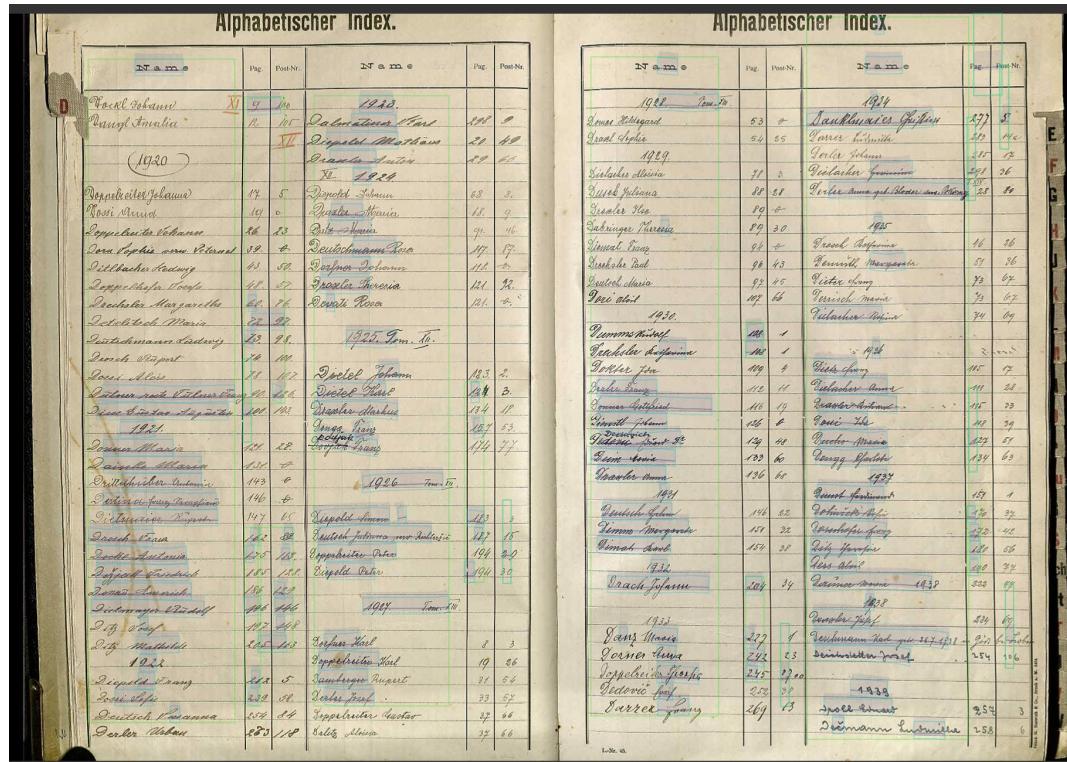
Lorenz



1918

# Off-the-shelf solutions don't always work

**Transkribus** (source:  
<https://readcoop.eu/transkribus/>)



The image shows two pages from an old ledger or account book. The pages are titled "Alphabetischer Index." and contain handwritten entries in German. The columns are labeled "Name," "Pag." (Page), and "Post-Nr." (Post Number). The entries are organized by date, indicated by handwritten dates like "1920," "1921," and "1922." The handwriting is cursive and varies in size and clarity. Some names are written in blue ink, while others are in black. The ledger appears to be a record of transactions or accounts for a period between 1920 and 1922.

Name	Pag.	Post-Nr.	Name	Pag.	Post-Nr.
Dickel Gabriele	11	100	1920		
Bauml. Amalia	11	101	Dalmatovics Paul	100	1
			Dimpfel Mathias	101	49
(1920)			Dimpfel Sophie	102	61
Zoppebauer Johanna	11	5	Dimpfel Sophie	103	3
Wohl Maria	10	6	Dimpfel Sophie	104	9
Zoppebauer Johanna	10	13	Dippe Anna	105	16
Sera. Sophie von Polenau	10	4	Deutschmann Rose	106	7
Schlauder Ludwig	11	37	Dorfner Lehre	107	5
Zoppebauer Sophie	11	57	Dorfner Sophie	108	32
Reichert Margaretha	11	76	Dorothy Rosa	109	5
Lechner Maria	12	32			
Gottschmann Ludwig	12	11	1921 Tom. 10.		
Drechsler August	13	10			
Löschner Alois	13	117	Dipfel Johann	110	2
Löschner und Löschner	13	118	Dipfel Paul	111	3
Drechsler August	13	119	Dipfel Sophie	112	11
Drechsler August	13	120	Ernst Hans	113	53
Drechsler Maria	13	121	Zoppebauer Sophie	114	77
Laville Maria	14	7			
Drauffelder Antonia	14	9	1922 Tom. 10.		
Löschner August	14	46			
St. Laurentius Kirche	14	61	Zoppebauer Anna	115	1
Drechsler Anna	14	88	Zoppebauer Sophie	116	15
Sander Adelheid	14	107	Zoppebauer Otto	117	20
Sander Amerilie	14	122	Zoppebauer Otto	118	30
Schönberger Elizab.	14	446			
Zoppebauer Sophie	14	448	1923 Tom. 10.		
Zoppebauer Mathilde	14	453	Zoppebauer Sophie	119	8
			Zoppebauer Sophie	120	26
Zoppebauer Sophie	14	5	Zumherren August	121	54
Zoppebauer Sophie	14	61	Zumherren August	122	56
Deutschmann Rosina	15	14	Zoppebauer Sophie	123	57
Gander Uebau	15	118	Zoppebauer Sophie	124	66
			Zoppebauer Sophie	125	66
			Zoppebauer Sophie	126	66
			Zoppebauer Sophie	127	66
			Zoppebauer Sophie	128	66
			Zoppebauer Sophie	129	66
			Zoppebauer Sophie	130	66
			Zoppebauer Sophie	131	66
			Zoppebauer Sophie	132	66
			Zoppebauer Sophie	133	66
			Zoppebauer Sophie	134	66
			Zoppebauer Sophie	135	66
			Zoppebauer Sophie	136	66
			Zoppebauer Sophie	137	66
			Zoppebauer Sophie	138	66
			Zoppebauer Sophie	139	66
			Zoppebauer Sophie	140	66
			Zoppebauer Sophie	141	66
			Zoppebauer Sophie	142	66
			Zoppebauer Sophie	143	66
			Zoppebauer Sophie	144	66
			Zoppebauer Sophie	145	66
			Zoppebauer Sophie	146	66
			Zoppebauer Sophie	147	66
			Zoppebauer Sophie	148	66
			Zoppebauer Sophie	149	66
			Zoppebauer Sophie	150	66
			Zoppebauer Sophie	151	66
			Zoppebauer Sophie	152	66
			Zoppebauer Sophie	153	66
			Zoppebauer Sophie	154	66
			Zoppebauer Sophie	155	66
			Zoppebauer Sophie	156	66
			Zoppebauer Sophie	157	66
			Zoppebauer Sophie	158	66
			Zoppebauer Sophie	159	66
			Zoppebauer Sophie	160	66
			Zoppebauer Sophie	161	66
			Zoppebauer Sophie	162	66
			Zoppebauer Sophie	163	66
			Zoppebauer Sophie	164	66
			Zoppebauer Sophie	165	66
			Zoppebauer Sophie	166	66
			Zoppebauer Sophie	167	66
			Zoppebauer Sophie	168	66
			Zoppebauer Sophie	169	66
			Zoppebauer Sophie	170	66
			Zoppebauer Sophie	171	66
			Zoppebauer Sophie	172	66
			Zoppebauer Sophie	173	66
			Zoppebauer Sophie	174	66
			Zoppebauer Sophie	175	66
			Zoppebauer Sophie	176	66
			Zoppebauer Sophie	177	66
			Zoppebauer Sophie	178	66
			Zoppebauer Sophie	179	66
			Zoppebauer Sophie	180	66
			Zoppebauer Sophie	181	66
			Zoppebauer Sophie	182	66
			Zoppebauer Sophie	183	66
			Zoppebauer Sophie	184	66
			Zoppebauer Sophie	185	66
			Zoppebauer Sophie	186	66
			Zoppebauer Sophie	187	66
			Zoppebauer Sophie	188	66
			Zoppebauer Sophie	189	66
			Zoppebauer Sophie	190	66
			Zoppebauer Sophie	191	66
			Zoppebauer Sophie	192	66
			Zoppebauer Sophie	193	66
			Zoppebauer Sophie	194	66
			Zoppebauer Sophie	195	66
			Zoppebauer Sophie	196	66
			Zoppebauer Sophie	197	66
			Zoppebauer Sophie	198	66
			Zoppebauer Sophie	199	66
			Zoppebauer Sophie	200	66
			Zoppebauer Sophie	201	66
			Zoppebauer Sophie	202	66
			Zoppebauer Sophie	203	66
			Zoppebauer Sophie	204	66
			Zoppebauer Sophie	205	66
			Zoppebauer Sophie	206	66
			Zoppebauer Sophie	207	66
			Zoppebauer Sophie	208	66
			Zoppebauer Sophie	209	66
			Zoppebauer Sophie	210	66
			Zoppebauer Sophie	211	66
			Zoppebauer Sophie	212	66
			Zoppebauer Sophie	213	66
			Zoppebauer Sophie	214	66
			Zoppebauer Sophie	215	66
			Zoppebauer Sophie	216	66
			Zoppebauer Sophie	217	66
			Zoppebauer Sophie	218	66
			Zoppebauer Sophie	219	66
			Zoppebauer Sophie	220	66
			Zoppebauer Sophie	221	66
			Zoppebauer Sophie	222	66
			Zoppebauer Sophie	223	66
			Zoppebauer Sophie	224	66
			Zoppebauer Sophie	225	66
			Zoppebauer Sophie	226	66
			Zoppebauer Sophie	227	66
			Zoppebauer Sophie	228	66
			Zoppebauer Sophie	229	66
			Zoppebauer Sophie	230	66
			Zoppebauer Sophie	231	66
			Zoppebauer Sophie	232	66
			Zoppebauer Sophie	233	66
			Zoppebauer Sophie	234	66
			Zoppebauer Sophie	235	66
			Zoppebauer Sophie	236	66
			Zoppebauer Sophie	237	66
			Zoppebauer Sophie	238	66
			Zoppebauer Sophie	239	66
			Zoppebauer Sophie	240	66
			Zoppebauer Sophie	241	66
			Zoppebauer Sophie	242	66
			Zoppebauer Sophie	243	66
			Zoppebauer Sophie	244	66
			Zoppebauer Sophie	245	66
			Zoppebauer Sophie	246	66
			Zoppebauer Sophie	247	66
			Zoppebauer Sophie	248	66
			Zoppebauer Sophie	249	66
			Zoppebauer Sophie	250	66
			Zoppebauer Sophie	251	66
			Zoppebauer Sophie	252	66
			Zoppebauer Sophie	253	66
			Zoppebauer Sophie	254	66
			Zoppebauer Sophie	255	66
			Zoppebauer Sophie	256	66
			Zoppebauer Sophie	257	66
			Zoppebauer Sophie	258	66
			Zoppebauer Sophie	259	66
			Zoppebauer Sophie	260	66
			Zoppebauer Sophie	261	66
			Zoppebauer Sophie	262	66
			Zoppebauer Sophie	263	66
			Zoppebauer Sophie	264	66
			Zoppebauer Sophie	265	66
			Zoppebauer Sophie	266	66
			Zoppebauer Sophie	267	66
			Zoppebauer Sophie	268	66
			Zoppebauer Sophie	269	66
			Zoppebauer Sophie	270	66
			Zoppebauer Sophie	271	66
			Zoppebauer Sophie	272	66
			Zoppebauer Sophie	273	66
			Zoppebauer Sophie	274	66
			Zoppebauer Sophie	275	66
			Zoppebauer Sophie	276	66
			Zoppebauer Sophie	277	66
			Zoppebauer Sophie	278	66
			Zoppebauer Sophie	279	66
			Zoppebauer Sophie	280	66
			Zoppebauer Sophie	281	66
			Zoppebauer Sophie	282	66
			Zoppebauer Sophie	283	66
			Zoppebauer Sophie	284	66
			Zoppebauer Sophie	285	66
			Zoppebauer Sophie	286	66
			Zoppebauer Sophie	287	66
			Zoppebauer Sophie	288	66
			Zoppebauer Sophie	289	66
			Zoppebauer Sophie	290	66
			Zoppebauer Sophie	291	66
			Zoppebauer Sophie	292	66
			Zoppebauer Sophie	293	66
			Zoppebauer Sophie	294	66
			Zoppebauer Sophie	295	66
			Zoppebauer Sophie	296	66
			Zoppebauer Sophie	297	66
			Zoppebauer Sophie	298	66
			Zoppebauer Sophie	299	66
			Zoppebauer Sophie	300	66
			Zoppebauer Sophie	301	66
			Zoppebauer Sophie	302	66
			Zoppebauer Sophie	303	66
			Zoppebauer Sophie	304	66
			Zoppebauer Sophie	305	66
			Zoppebauer Sophie	306	66
			Zoppebauer Sophie	307	66
			Zoppebauer Sophie	308	66
			Zoppebauer Sophie	309	66
			Zoppebauer Sophie	310	66
			Zoppebauer Sophie	311	66
			Zoppebauer Sophie	312	66
			Zoppebauer Sophie	313	66
			Zoppebauer Sophie	314	66
			Zoppebauer Sophie	315	66
			Zoppebauer Sophie	316	66
			Zoppebauer Sophie	317	66
			Zoppebauer Sophie	318	66
			Zoppebauer Sophie	319	66
			Zoppebauer Sophie	320	66
			Zoppebauer Sophie	321	66
			Zoppebauer Sophie	322	66
			Zoppebauer Sophie	323	66
			Zoppebauer Sophie	324	66
			Zoppebauer Sophie	325	66
			Zoppebauer Sophie	326	66
			Zoppebauer Sophie	327	66
			Zoppebauer Sophie	328	66
			Zoppebauer Sophie	329	66
			Zoppebauer Sophie	330	66
			Zoppebauer Sophie	331	66
			Zoppebauer Sophie	332	66
			Zoppebauer Sophie	333	66
			Zoppebauer Sophie	334	66
			Zoppebauer Sophie	335	66
			Zoppebauer Sophie	336	66
			Zoppebauer Sophie	337	66
			Zoppebauer Sophie	338	66
			Zoppebauer Sophie	339	66
			Zoppebauer Sophie	340	66
			Zoppebauer Sophie	341	66
			Zoppebauer Sophie	342	66
			Zoppebauer Sophie	343	66
			Zoppebauer Sophie	344	66
			Zoppebauer Sophie	345	66
			Zoppebauer Sophie	346	66
			Zoppebauer Sophie	347	66
			Zoppebauer Sophie	348	66
			Zoppebauer Sophie	349	66
			Zoppebauer Sophie	350	66
			Zoppebauer Sophie	351	66
			Zoppebauer Sophie	352	66
			Zoppebauer Sophie	353	66
			Zoppebauer Sophie	354	66
			Zoppebauer Sophie	355	66
			Zoppebauer Sophie	356	66
			Zoppebauer Sophie	357	66
			Zoppebauer Sophie	358	66
			Zoppebauer Sophie	359	66
			Zoppebauer Sophie	360	66
			Zoppebauer Sophie	361	66
			Zoppebauer Sophie	362	66
			Zoppebauer Sophie	363	66
			Zoppebauer Sophie	364	66
			Zoppebauer Sophie	365	66
			Zoppebauer Sophie	366	66
			Zoppebauer Sophie	367	66
			Zoppebauer Sophie	368	66
			Zoppebauer Sophie	369	66
			Zoppebauer Sophie	370</td	

# Challenges

## Images

- Tables are not uniform in structure (variation in fields/columns), owing to the range of periods and geographic locations the dataset comes from
- Handwriting styles vary due to differences in human-writers and periods
- Noise present in the form of ink blots, parchment texture, or general wear and tear of the physical documents

MATRICULA.								COPULATORUM.			
Nrs. cur.	Annus et dies copulationis	S P O X S I .			S t a t u s			Nomen et Conditio testium	Nomen et Officium Copulatis	nam ponsibus ut vel dispositi sati in bancis vel aliquo in- pudenter?	Observationes
		Nomen et Conditio sponsi et sponsae ac parentum	latus originis et domus corundem	Religo	Aetas	coelos	vibus				
8.	1854	Johann Seitz Anna magdalena Hans Seitz Stephanus Seitz Johanna Barbara Barbara Gottlieb Gottlieb Barbara Karl Gottlieb Gottlieb Barbara Felix Gottlieb Gottlieb Barbara	Steinbach Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 23. 23. 23. 23.	1. 1. 1. 1. 1.	— — — — —	Alexander Hanella arela Hanella magdalena Barbara Barbara Barbara	Fratres Barbara Barbara	Piemont Piemont	Tunc postmodum oblationem nuptiale Piemont etiam dicitur etiam pro Piemont aliisque etiam invenimus inde cum eis omnes portas resolutas de omnibus colonia in Regione Salisburgensis Piemont dicitur nuptiale.
	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —				
9.	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —	Johann Kirchmayer colonus maria spes in Götzendorf	Idem. Idem.	Piemont Piemont	Tunc postmodum oblationem nuptiale Piemont etiam dicitur etiam pro Piemont aliisque etiam invenimus inde cum eis omnes portas resolutas de omnibus colonia in Regione Salisburgensis Piemont dicitur nuptiale.
	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —				
10.	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —	Bartholomäus Litsch colonus maria spes in Götzendorf	Idem. Idem.	Piemont Piemont	Tunc postmodum oblationem nuptiale Piemont etiam dicitur etiam pro Piemont aliisque etiam invenimus inde cum eis omnes portas resolutas de omnibus colonia in Regione Salisburgensis Piemont dicitur nuptiale.
	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —				
11.	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —	Jacobus Lang colonus maria spes in Götzendorf	Idem. Idem.	Piemont Piemont	Tunc postmodum oblationem nuptiale Piemont etiam dicitur etiam pro Piemont aliisque etiam invenimus inde cum eis omnes portas resolutas de omnibus colonia in Regione Salisburgensis Piemont dicitur nuptiale.
	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —				
12.	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —	Petrus Pfeiffer colonus maria spes in Götzendorf	Piemont Piemont	Piemont Piemont	Tunc postmodum oblationem nuptiale Piemont etiam dicitur etiam pro Piemont aliisque etiam invenimus inde cum eis omnes portas resolutas de omnibus colonia in Regione Salisburgensis Piemont dicitur nuptiale.
	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —				
13.	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —	Petrus Schell colonus maria spes in Götzendorf	Piemont Piemont	Piemont Piemont	Tunc postmodum oblationem nuptiale Piemont etiam dicitur etiam pro Piemont aliisque etiam invenimus inde cum eis omnes portas resolutas de omnibus colonia in Regione Salisburgensis Piemont dicitur nuptiale.
	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —				
14.	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —	Petrus Schell colonus maria spes in Götzendorf	Piemont Piemont	Piemont Piemont	Tunc postmodum oblationem nuptiale Piemont etiam dicitur etiam pro Piemont aliisque etiam invenimus inde cum eis omnes portas resolutas de omnibus colonia in Regione Salisburgensis Piemont dicitur nuptiale.
	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —				
15.	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —	Petrus Schell colonus maria spes in Götzendorf	Piemont Piemont	Piemont Piemont	Tunc postmodum oblationem nuptiale Piemont etiam dicitur etiam pro Piemont aliisque etiam invenimus inde cum eis omnes portas resolutas de omnibus colonia in Regione Salisburgensis Piemont dicitur nuptiale.
	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —				
16.	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —	Petrus Schell colonus maria spes in Götzendorf	Piemont Piemont	Piemont Piemont	Tunc postmodum oblationem nuptiale Piemont etiam dicitur etiam pro Piemont aliisque etiam invenimus inde cum eis omnes portas resolutas de omnibus colonia in Regione Salisburgensis Piemont dicitur nuptiale.
	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —				
17.	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —	Petrus Schell colonus maria spes in Götzendorf	Piemont Piemont	Piemont Piemont	Tunc postmodum oblationem nuptiale Piemont etiam dicitur etiam pro Piemont aliisque etiam invenimus inde cum eis omnes portas resolutas de omnibus colonia in Regione Salisburgensis Piemont dicitur nuptiale.
	1854	Wenzel Gottlieb Barbara Wenzel Gottlieb Barbara Anna Maria Barbara Barbara Felix Gottlieb Barbara Barbara Felix Gottlieb Barbara Barbara	Götzendorf Götzendorf Götzendorf Götzendorf Götzendorf	Religio Cathol.	23. 24. 24. 24. 24.	1. — 1. — 1.	— — — — —				
Liquidus Anno Domini 1855.											

A parish register from Austria, 1855 (source: <https://data.matricula-online.eu/>)

# Challenges

## Images

- Poor contrast
- Overlapping text across columns
- Disregard any miscellaneous scribbles/notes

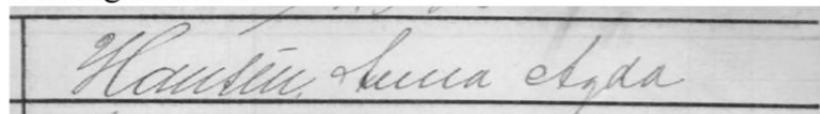
Post-Nr.	Jahr 1913 Monat: Geburts- und Tag des Geburts- Taufe	Ort Haus- Nr.	Namens- Kinder	Geschlecht Lebendig geborenen Kinder dabei verstorben nachgeboren oder gestorben	Der Religion	Eltern		Der Paten Namens und Stand	Name Vater	Name Mutter	Name und Stand	Täufer	Hebamme
						Eltern	Der Paten						
376	Oktober 1. 1913	Blumau 5	Leggolt Johanna	Lebendig geboren am 26.10.1913 gestorben am 26.10.1913	1	1	1	Sister Pommer Brautfeier der Postleitz geboren am 22.8.1888 am 9.5.1913	Julia	geb. Kastner	Maria Krebs geb. in Leibnitz verheirathet mit Gottlieb eines Pommers Leggolt katholisch Gottgott	Franz Schwagler Karl Fischer	Johanna Krebs Karl Fischer
376	November 2. 1913	Riedhofdorf	Karl	Lebendig geboren am 23.11.1913 gestorben am 23.11.1913	1	1	1	Julia Schöberl Brautfeier der Postleitz geboren am 26.10.1873 verheirathet mit Anton Schöberl aus Leibnitz katholisch Gottgott	Julia Schöberl	geb. Kastner	Franz Oberndorfer Brautfeier der Postleitz geboren am 26.10.1873 verheirathet mit Anton Schöberl aus Leibnitz katholisch Gottgott	Franz Oberndorfer Karl Fischer	Julia Schöberl Karl Fischer
376	November 9. 1913	Reith 5	Johanna	Lebendig geboren am 10.11.1913 gestorben am 10.11.1913	1	1	1	Julia Krenic Brautfeier der Postleitz geboren am 01.01.1890 verheirathet mit Karl Krenic aus Leibnitz katholisch Gottgott	Julia Krenic	geb. Schmid	Franz Krenic Karl Fischer	Franz Krenic Karl Fischer	Julia Krenic Karl Fischer
376	November 11. 1913	Reithof 65	Franz	Lebendig geboren am 12.11.1913 gestorben am 12.11.1913	1	1	1	Franz Postleitz Brautfeier der Postleitz geboren am 09.03.1882	Julia Kastner	geb. Kastner	Franz Kastner Karl Kastner	Franz Kastner Karl Kastner	Franz Kastner Karl Kastner
376	November 14. 1913	Riedhofdorf	Wadim	Lebendig geboren am 14.11.1913	1	1	1	Julia Kastner Brautfeier der Postleitz geboren am 09.03.1882 verheirathet mit Franz Kastner katholisch Gottgott	Julia Kastner	geb. Kastner	Franz Kastner Karl Kastner	Franz Kastner Karl Kastner	Julia Kastner Karl Kastner

A parish birth register from Austria (source: <https://data.matricula-online.eu/>)

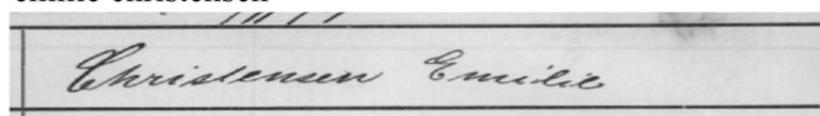
# Using custom-made deep learning models for OCR

- Fine-tuning deep neural networks requires large amounts of data (>1 million images with labels)
- HAndwritten NAme (HANA) database:
  - Danish Census data, 1890-1923
  - Handwriting looks similar to that in our parish records
  - **~1.1 million** handwritten names
  - Accurate labels and clean images

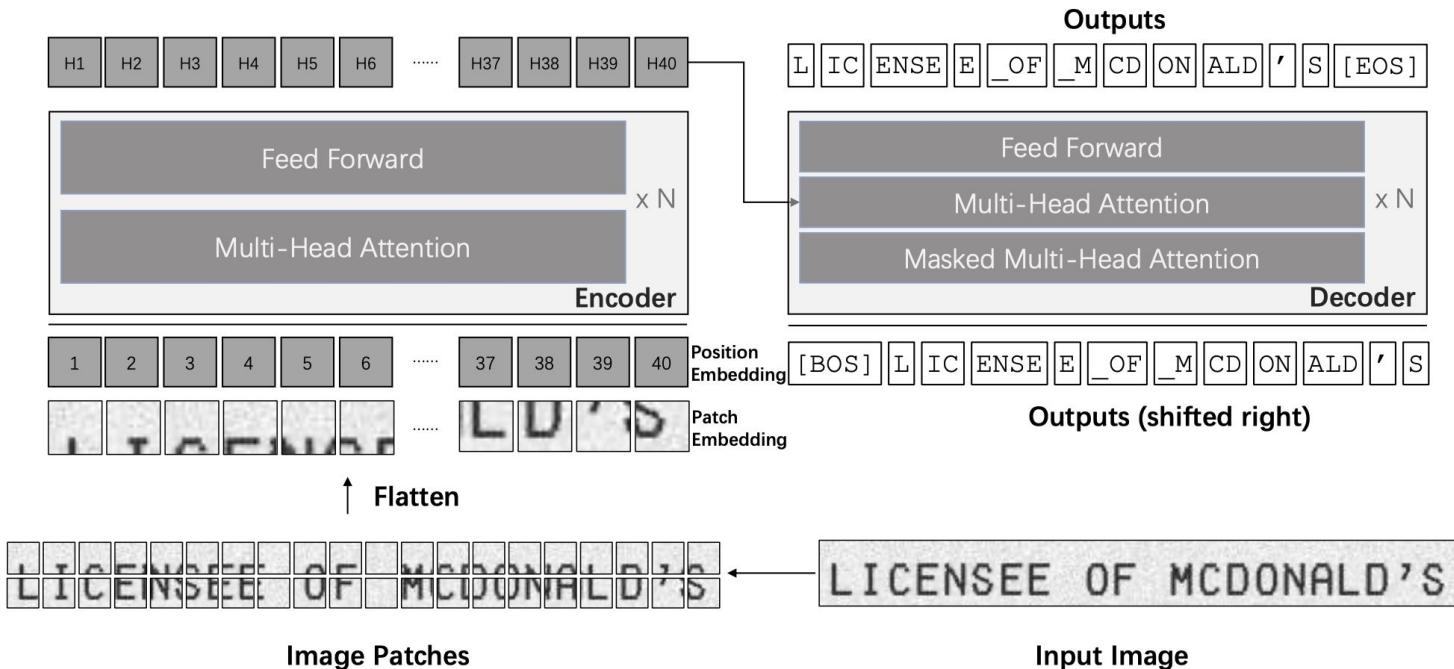
anna agda hansen



emilie christensen



# TrOCR: Transformer-based OCR model



# Resources utilised

- Monash M3 MASSIVE high-performance computing (HPC) cluster
  - Provides GPU nodes for running DL models
  - Large temporary storage and memory



- Amazon Web Services (AWS) (Commercial Cloud)
  - Lambda functions to run web-scrapers to fetch images from Matricula website
  - S3 buckets to store image repositories scraped from the web



Thank you!